Audio Engineering Society

Convention Paper

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

Development and Evaluation of an Interface with Four-Finger Pitch Selection

Henrik von Coler[1], Gabriel Treindl[1], Hauke Egermann[2], and Stefan Weinzierl[1]

[1]*Audio Communication Group, TU-Berlin*
[2]*Department of Music , University of York*

Correspondence should be addressed to Henrik von Coler (voncoler@tu-berlin.de)

ABSTRACT

In this paper an interface for electronic musical instruments is presented, which is primarily designed for playing monophonic synthesizers. The hand-held device allows the pitch selection with one hand, using four valve-like metal mechanics and three octave switches. Note events are triggered with a wooden excitation pad, operated with the second hand. Another feature is the advanced aftertouch of the four mechanics and the pad, which enables expressive playing. In a user experiment, the controller is compared to a classic MIDI keyboard, regarding the time needed for responding to simple visual stimuli and the mean error rate produced in that task. The results show no significant difference in the response time but a higher error rate for the novel interface for untrained users. Outcome of this work is a list of necessary improvements, as well as a plan for further experiments.

Introduction

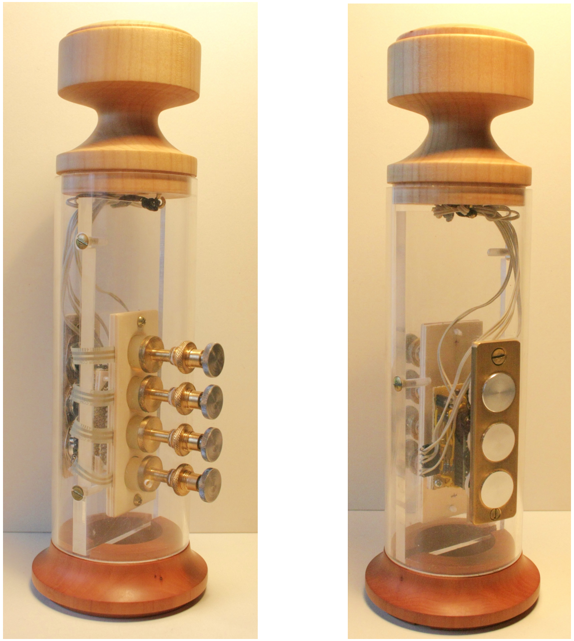
The motivation for developing this novel interface, is to allow a more expressive control for electronic melody instruments. The hand-held controller, shown in Figure 1, uses one hand for pitch selection, based on binary combination of four fingers, and the second hand for note triggering, dynamics and timbre control. It revives the concept of aftertouch, as known from traditional and recent synthesizers, by making the four valve-like mecahnics for pitch selection sensitive to pressure, using force sensitive resistors (FSR). The combination of pitch selection and trigger allows the use with percussive synthesis as well as pads and drones and the flexible application of different articulation styles.

The MIDI-Keyboard became the most widespread and popular interface for good reasons, due to its versatility and intuitivity. Monophonic synthesizers, however, benefit from a paradigm, where pitch selection and excitation are decoupled. This allows more control over the tone, in the case of articulation styles and modulations. It is known from classical, mechanical instruments that mainly monophonic instruments allow expressive manipulations as vibrato and glissando [1].

The BinBong was developed in an interdisciplinary team, consisting of engineers from the Audio Communication Group at TU Berlin and scientists from the SIM[1], respectively the restoration workshop for musical instruments located there.

The remainder of this paper is organized as follows: Section 2 briefly discusses the relevant principles of control, which form the foundation of the interface.

[1]Staatliches Institut für Musikforschung, Berlin



(a) Front view, showing the valve mechanics

(b) Rear view, showing the octave switches

Fig. 1: Front (a) and rear (b) view of the BINBONG

Section 3 summarizes the techical realization, considering hardware design decisions and programming aspects. The experiment for comparing the response time and error rate of the device with a generic MIDI keyboard is presented in Section 4, alongside the evaluation and discussion of the results. Final aspects are treated in the Conclusion in Section 5.

Control Principles

Pitch Selection

The pitch selection is carried out using combinations of four fingers operating the valve-mechanics. Each finger rests on the respective mechanic, minimizing hand movements in play. Binary combinations allow $2^4 - 1 = 15$ different pitches to be played, since the case of no pressed mechanic is ignored. The initial mapping of combinations to one octave which is also evaluated in the experiment, is based on binary[2] counting, as shown in Table 1. Different mappings, for example based on gray codes [3], are possible and maybe beneficial for several reasons. However, the chosen straight-forward mapping is considered to be intuitive and self-explanatory. Since the experiments at this stage aim at a general response time and error rate, the mapping is not considered to have an impact.

[2]This leads to the term 'BIN' in the working title.

Table 1: Binary Code Mapping - *F5 = little finger, F4 = ring finger, F3 = middle finger, F2 = index*

F5	F4	F3	F2	
○	○	○	•	C
○	○	•	○	C#
○	○	•	•	D
○	•	○	○	D #
○	•	○	•	E
○	•	•	○	F
○	•	•	•	F #
•	○	○	○	G
•	○	○	•	G #
•	○	•	○	A
•	○	•	•	A #
•	•	○	○	B
•	•	○	•	C'

Three additional octave buttons are located on the opposite side of the valve mechanics, as shown in Figure 1b. Operated with the thumb, they allow the selection of up to six octaves, allowing the joint activation of adjacent buttons. The experiment presented in this paper makes use of only one octave and thus neglects the octave buttons, completely.

Modulation

Modulations of sound parameters are crucial for an expressive musical performance, especially periodic modulations of pitch, amplitude and spectral shape [4]. Providing means for expressive modulations is thus a key claim in designing interfaces for this type of application. Different concepts of pitch modulation in DMIs have been investigated by Marshall et. al. [5].

The BINBONG principle, which may be regarded as a further development of the classic aftertouch, is inspired by observing how vibrato is applied in most physical instruments. It is rooted on the hypothesis that vibrato is induced by applying general muscular tension [6], which is then translated to a semi-voluntary tremor of a frequency of $5 - 12$Hz.

Note Triggering

The BinBong design supplies a dedicated pad for triggering note events. It can be played with the thumb or the palm. Pitch selection and 'excitation' are thus decoupled, which creates more degrees of freedom in expressive performance and allows the intuitive application of different articulation styles. This decoupling is also typical for classical physical melody instruments (Strings, Woodwinds, Brass), whereas most polyphonic instruments (Piano, Organ) combine pitch selection and excitation.

Technical Realization

Housing

The realization of the prototype was mainly facilitated through a collaboration between engineers, resposible for the electronics, and an instrument builder, who took over the mechanical development. Starting point was the request to place at least four valve-like mechanics for pitch selection in a device which can be held in a single hand, by both left-handed and right-handed users. Sketches for possible prototypes which were considered are shown in Figure 2. Figure 2b shows the first design incorporating the excitation pad.

The housing, shown in Figure 3 as a cross section, consists of an acrylic cylinder[3] with a diameter of 70 mm and a length of 180 mm. An acrylic platform through the longitudinal section holds the electronics and the force sensitive resistors (FSRs) for the valve mechanics. The wooden base, which offers cavities for the USB cable, allows a stable stand of the device.

Valve Mechanics

The four valve mechanics, shown in Figure 4 as close-up, represent electrical keys with a highly sensitive aftertouch. The latter is realized using force sensitive resistors (FSR). FSRs are the most widely used sensors in the design of DMI, justified by the fact that they result in highest preference in user tests [5]. Custom-compounded and molded soft silicone cushions are situated between the plunger and the FSR. They do not only act as reset force but in this way the mechanics have a soft action point and allow a movement of of about 5 mm. This enables a finer dosing of the force applied and thus a more precise control.

[3]This accounts for the 'Bong' in the working title.

Octave Switches

Three piezoelectric pushbuttons serve as octave switches. This type of switch feels like a capacitive sensor, since there is no pushing in, yet they show less inertia. The switches are almost planar (see Figure 3), making them palpable but not distracting.

Excitation Pad

The wooden excitation pad is bedded on four silicone cushions, which can be seen in Figure 5. A FSR is located under each cushion. The relative force of these sensors is averaged and mapped to the note velocity at this stage. Additionally, two main axis of the pad are obtained and used for sound control.

Electronics and Programming

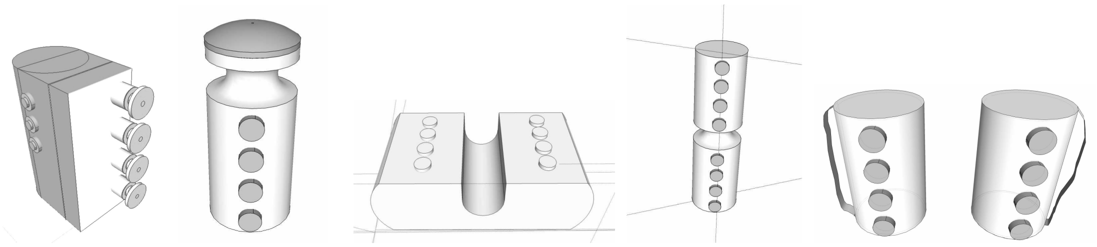
The complete processing is realized on a Teensy 3.1 board, which is connected to the sensors.For the implementation in this stage, the microcontroller evaluates all sensor inputs and sends them to a Pure Data patch for further processing and synthesis. The synthesis algorithm consists of a simple subractive approach for the evaluation.

The force applied to the four valve-mechanics is averaged, resulting in one single aftertouch parameter. It is most likely not feasible to use each sensor, separately. In order to avoid glitches when pressing combinations, a safety window of 46 ms is used for waiting for additional mechanics to be pressed after one has been activated. This value has been tuned heuristically for beginners and it can be decreased after a certain period of training.

Evaluation study

In general, the quality and usability of interfaces for musical instruments may not be investigated quantitatively. These aspects always depend on the context of application strongly and might have completely different dimensions, depending on the musical application. However, the evaluation problem for new musical interfaces [8] is of interest for the research community and needs further development.

For melody instruments in conventional popular music we suggest to split the evaluation problem into several dimensions. The basic ones are considered to be



(a) Initial design **(b)** BinBong **(c)** Palm design **(d)** Vertical design **(e)** Decoupled design

Fig. 2: Different design studies for devices with the valve mechanics

response time and error rate in fast play, accuracy in frequency selection and modulation capabilities. The experiment presented in this paper is designed as a method for comparing only the response time and error rate and aims at generating a generalized procedure, applicable to other interfaces. This may be regarded a first validation of the principle, in order to pave the way for experiment dealing with the remaining aspects.

Since the MIDI piano is still the predominant interface, it is considered a valid reference for the evaluation. Therefore, a conventional keyboard was used as reference interface. For the experiment, both interfaces were reduced to the range of a single octave.

Due to the many differences between the two interfaces, divergent performance results would be difficult to explain. Thus, several additional control conditions were added, in order to understand any emerging differences in instrument performance:

a) Two different versions of the controller are evaluated: Since pitch selection and triggering are decoupled in the original setup, the excitation pad has to be used to play the note. In a modified version for this evaluation experiment, the note was played directly when the valve mechanics are pressed. Comparing these two settings allowed to determine whether the additional excitation pad would reduce playing errors, as with it being activated, only complete finger combinations could be executed separately by the user.

b) Playing all twelve notes of one octave requires finger combinations on the controller and moving the finger and hand in the horizontal dimension on the keyboard. Any emergent differences between keyboard and BinBong performance could then be due to the use of different finger combinations, the lack of horizontal movements in the BinBong, or the use of different mechanical buttons. We therefore added for both interfaces a control condition with just 4 notes to be played. This results in a fixed fingering for the keyboard and avoids finger combinations for the controller.

Accordingly, we created two experimental factors: three different interface configurations and two different task versions (12 different diatonic pitches vs. just 4). Each participant completed each of the resulting six experimental conditions listed in Table 2.

Table 2: Overview of experimental conditions

	Task difficulty: Number of different pitches to play	
Interface	4	12
BinBong with excitation pad (BB_{pad})	Condition 1	Condition 4
BinBong without excitation pad (BB_{nopad})	Condition 2	Condition 5
Keyboard (KB)	Condition 3	Condition 6

Subjects

Out of the 20 participants were 18 male, 45% were students and 55% full-time employees. Their age was between 23 and 34 years (mean = 28.84, SD = 3.20). 20% of the participants indicated that they preferred to play the interfaces with their left hand. All of them had played a musical instrument before and half of them received piano lessons for more than one year and practiced piano more than half an hour per day during this period. None of the participants identified themselves as a professional musician and none of them played a valve instrument.

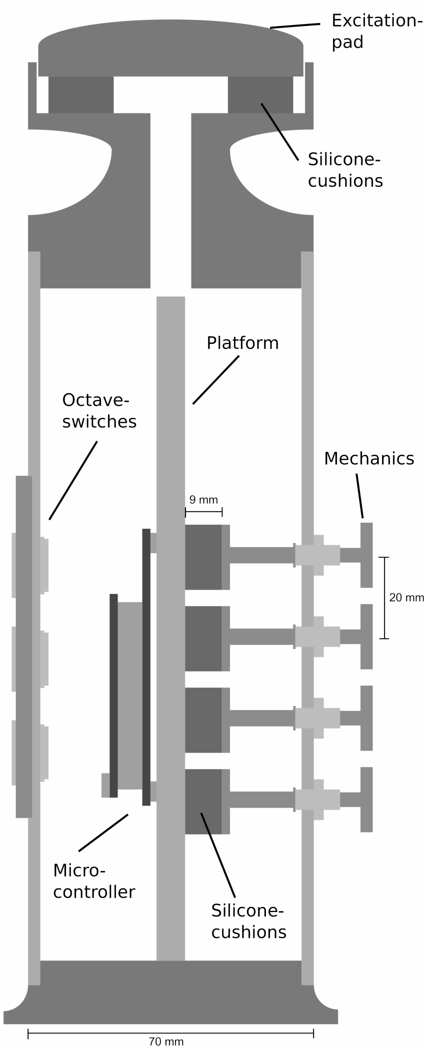


Fig. 3: Cross section showing housing and mechanics

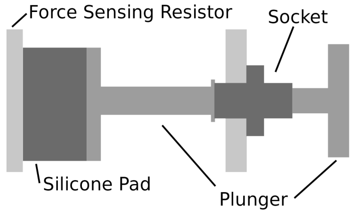


Fig. 4: Valve mechanic close-up

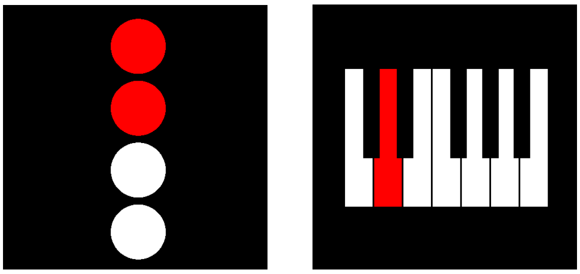
Fig. 5: Top view with removed excitation pad, showing the four silicone cushions with FSRs below

Experimental Setup

The complete test procedure was implemented in Pure Data, including the presentation of the visual stimuli, the sound synthesis for auditive feedback, as well as the recording of the responses. A computer screen was used to present pitches as visual stimuli, generated with the GEM library [9], as shown in Figure 6. For the keyboard interface, only one octave was displayed and for the BinBong interface four vertically arranged circles were shown on the screen. Keys which had to be pressed to play the required pitch were marked red. The controller was connected to the computer via USB, directly, whereas the keyboard was connected to the MIDI-input of the sound card. Activated keys were displayed on the screen in green, in order to provide additional visual feedback. An auditive feedback was generated in Pure Data, using a simple subtractive synthesis algorithm with fixed parameters and without velocity mapping. Depending on the task, random permutated sequences of four or twelve different pitches were generated in advance for each subject and condition.

Procedure

In a within subjects design, every participant completed all six conditions from Table 2 in random order. Each condition included a training phase with 24 trials, respectively 24 notes to be played, and a test run with 60 test trials. Participants were instructed to play each note as quickly as possible, when the color of the required keys turned red, trying to minimize the number of wrong responses. After the correct pitch was played, the stimulus for the next note was presented as soon as the keys were released. When all 24 trials of the training phase and 60 trials of test run were completed, the next condition was presented.



(a) Instruction to press index and ring finger on the BINBONG

(b) Instructions on the screen, as used in the experiment

Fig. 6: Instructions on the screen, as used in the experiment

Measurements

The following data was stored for each note played by the participants: The MIDI note numbers of the played and presented notes, the current timer value and the trial number. The time from displaying the visual stimuli until receiving the MIDI Note-On message is measured in milliseconds. Subtracting the average MIDI latency [10] of the controller resulted in the response time of each participant and trial. Subsequently, we calculated for each participant and condition a) the mean response time to press the correct note and b) the mean error rate (both across all 60 test trials).

Results and Discussion

The significance of experimental factors was evaluated through estimation of two linear models using the SPSS Mixed procedure (one for mean response time and another one for mean errors as dependent variable). Model fit indices of AIC and BIC indicated that for both models, a covariance structure with compound symmetry and heterogeneous variances fitted best. That way, we modelled the dependency in residuals in repeated observations. Figure 7a presents box plots for response time and Figure 7b respective box plots for the errors per condition, separated by experimental factors.

Errors

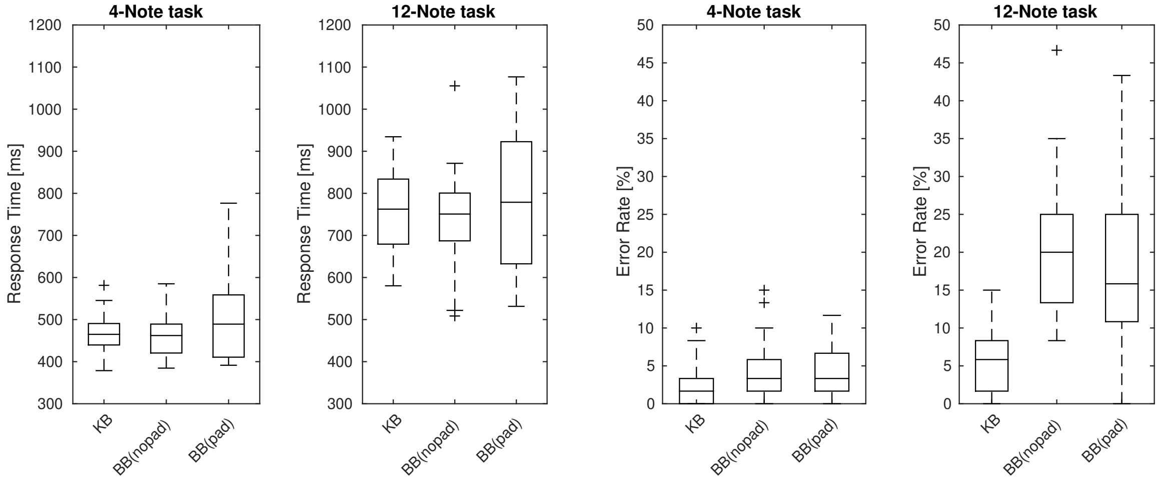
Tab.3 presents coefficient estimates and associated inferential statistics for all experimental conditions and their interactions predicting the number of errors per condition, respectively. These indicate that both versions of the BINBONG (with and without the excitation pad activated) resulted in higher errors as compared with the keyboard controller (used as a reference category). Furthermore, increasing the complexity of the task by presenting 12 different pitches instead of only 4 also increased the number of wrong notes played by participants. This difference also became larger for the two BINBONG versions, as indicated by the significant interaction terms. These findings indicate, that in general, the BINBONG was more difficult to play accurately than the keyboard. However, the significant interaction indicates that this was most likely due to the increased difficulty of playing finger combinations in the 12 pitch conditions. In the 4 pitch condition without finger combinations, the interfaces are much less different from each other. Furthermore, introducing the excitation pad, helped to decrease the number of errors, as participants had the possibility press it only once all fingers were in their correct position (however, this difference was not significant as indicated by a paired t-Test).

Table 3: Coefficient estimates from predicting error frequency through linear model: BB_{pad} = BINBONG with excitation pad vs. Keyboard, BB_{nopad} = BINBONG without excitation pad vs. Keyboard, 12P: 12 pitches vs. 4 pitches, BB_{pad}x12P = BINBONG with excitation pad, BB_{nopad}x12P = BINBONG without excitation pad

Predictor	b estimate	df	t	p
Intercept	2.3	20.2	3.6	0.002
BB_{pad}	1.9	35.0	2.2	0.032
BB_{nopad}	2.4	32.2	2.4	0.022
12P	3.1	31.3	2.9	0.007
sBB_{pad}x12P	10.3	27.7	4.0	0
BB_{nopad}x12P	12.7	34.1	5.3	0

Response Time

Table 4 presents coefficient estimates and associated inferential statistics for all experimental conditions and their interactions predicting the response times. Response time was slightly longer for the BINBONG version with the excitation pad, indicated by a non-significant trend ($p<.10$). However, the BINBONG without the excitation pad activated did not lead to slower response times than the keyboard. Again, increasing



(a) Response time **(b)** Error Rate

Fig. 7: Box plots with the results for response time (a) and error rate (b)

the task complexity (playing 12 vs. only 4 different pitches), lead to a significant increase in response time. However, there are no significant interaction terms between the experimental factors type of interface and task difficulty.

Table 4: Coefficient estimates from predicting response time through linear model: BB_{pad} = BINBONG with excitation pad vs. Keyboard, BB_{nopad} = BINBONG without excitation pad vs. Keyboard, 12P: 12 pitches vs. 4 pitches, BB_{pad}x12P = BINBONG with excitation pad, BB_{nopad}x12P = BINBONG without excitation pad

Predictor	b estimate	df	t	p
Intercept	469.3	20.6	45.5	0
BB_{pad}	38.5	24.2	1.9	0.069
BB_{nopad}	-9.4	43.8	-0.9	0.384
12P	288.7	22.3	14.4	0
BB_{pad}x12P	5.6	36.1	0.1	0.895
BB_{nopad}x12P	63.9	19.9	0.9	0.395

The results of this experimental evaluation indicate that the BINBONG can be operated with similar response times as the conventional keyboard interface. However, when complex finger combinations were required in order to play 12 different pitches, the keyboard interface without finger combinations was easier to use. Adding the excitation pad to the BINBONG helped to reduce the number of errors, yet not significantly, and lead to a slightly increased response time. All participants had experience with using the conventional keyboard to control pitch. However, no participant in this study was experienced in playing an instrument with finger combinations, as for example woodwinds or brass instruments. This might introduce a potential bias in these findings: Longer training in using this novel controller could create more positive results. Also, repeating the experiment with a range of more than one octave would shift the performance.

Conclusion

We regard the BINBONG a promising step towards a simple interface with expressive capabilities for electronic melody instruments. The valve mechanics with the incorporated aftertouch are a convincing alternative for standard keys and their further use will be promoted. Participants of the experiment also filled out a survey which evaluated the device. The overall response was positive and also revealed necessary improvements, some of them already detected by the development team.

As a result of the first tests, several changes are recently beeing implemented in the next stage of development. Most important is equipping the device with a wireless comunication, as well as changes in ergonmics, respectively the relative positions of the control elements and the tube diameter, which needs to be decreased. The octave switches need to be relocated and improved, the piezo approach seems very useful.

The result of the experiment with untrained users indicates that the novel controller can be operated with response times comparable to a MIDI keyboard, yet with higher error rates. Future work will focus on further experiments, once the necessary changes in hardware are realized. These experiments aim at finding best solutions for the programming and the mapping. This includes the task of finding an optimal set of 4-finger combinations for pitch mapping and the interaction with the octave switches. Upcoming tests will also explore the use in more complex musical tasks, including the learning carve.

Recently, the device is being integrated into a real-time application for spectral modeling synthesis. Thus, different mapping strategies of sensor data on sonic parameters can be investigated in user experiments.

Acknowledgments

We especially value and emphasize the work which Dr. Tom Lerch from the SIM has contributed to this project. As a trained instrument builder, he was not only able to manage the production of the prototype, but also to influence the design, according to his long-lasting experience in the work with musical instruments.

References

[1] Levitin, D. J., McAdams, S., and Adams, R. L., "Control Parameters for Musical Instruments: A Foundation for New Mappings of Gesture to Sound," *Org. Sound*, 7(2), pp. 171–189, 2002, ISSN 1355-7718, doi:10.1017/S135577180200208X.

[2] Paradiso, J. A., "Electronic music: new ways to play," *IEEE Spectrum*, 34(12), pp. 18–30, 1997, ISSN 0018-9235, doi:10.1109/6.642965.

[3] Beauregard, G., *Rethinking the Design of Wind Controllers*, Dartmouth College, 1991.

[4] Verfaille, V., Guastavino, C., and Depalle, P., "Perceptual Evaluation of Vibrato Models," *Proceedings of the Conference on Interdisciplinary Musicology (CIM05)*, 2005.

[5] Marshall, M. T., Hartshorn, M., Wanderley, M. M., and Levitin, D. J., "Sensor Choice for Parameter Modulations in Digital Musical Instruments: Empirical Evidence from Pitch Modulation," *Journal of New Music Research*, 38(3), pp. 241–253, 2009, doi:10.1080/09298210903085865.

[6] Fletcher, N. H., "Vibrato in music - physics and psychophysics," *Proceedings of the International Symposium on Music Acoustics*, 2010.

[7] Arfib, D., Couturier, J.-M., and Kessous, L., "Expressiveness and digital musical instrument design," 34(1), pp. 125 – 136, 2005.

[8] Barbosa, J., Malloch, J., Wanderley, M., and Huot, S., "What does 'Evaluation' mean for the NIME community?" in E. Berdahl and J. Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, ????

[9] Danks, M., "Real-time image and video processing in gem," in *Proceedings of the International Computer Music Conference*, pp. 220–223, 1997.

[10] Wright, M., Cassidy, R. J., and Zbyszyński, M. F., "Audio and Gesture Latency Measurements on Linux and OSX," pp. 423–429, Miami, FL, USA, 2004.